eLIFE
elifesciences.org

Figures and figure supplements

Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels

Lizbeth Sayavedra, et al.

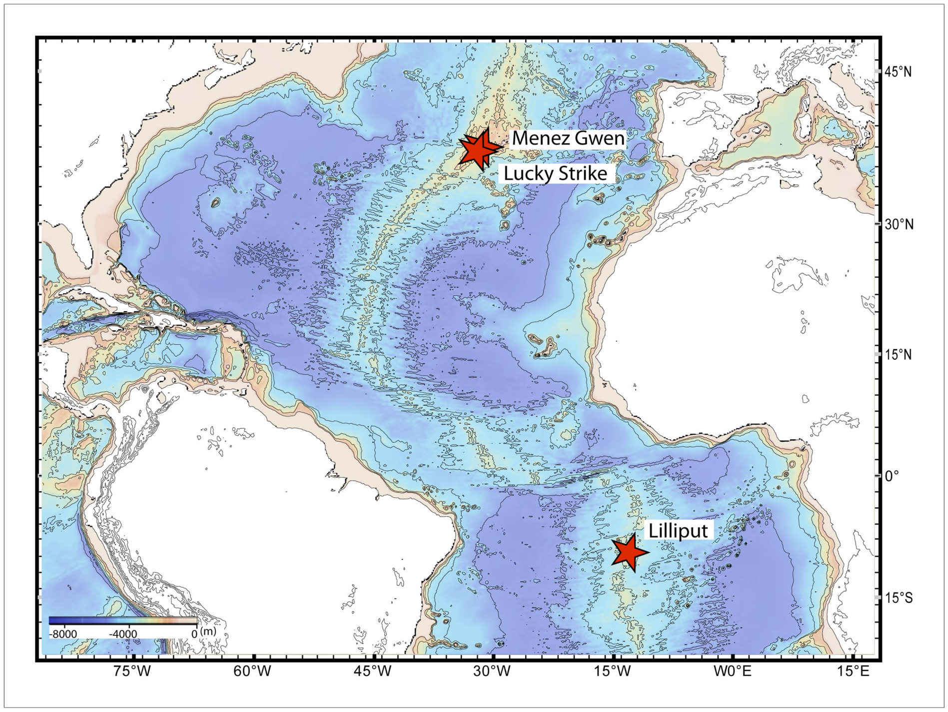

Figure 1. Sampling sites. Map showing the sampling sites of *Bathymodiolus* mussels at hydrothermal vents along the Mid-Atlantic Ridge (red stars). *B.* sp. is found at Lilliput (BspSym), *Bathymodiolus azoricus* at Menez Gwen (BazSymA and BazSymB) and Lucky Strike. The details of the sampling sites are described in ***Supplementary file 1E***. The map was produced with GeoMapApp 3.3.

DOI: 10.7554/eLife.07966.003

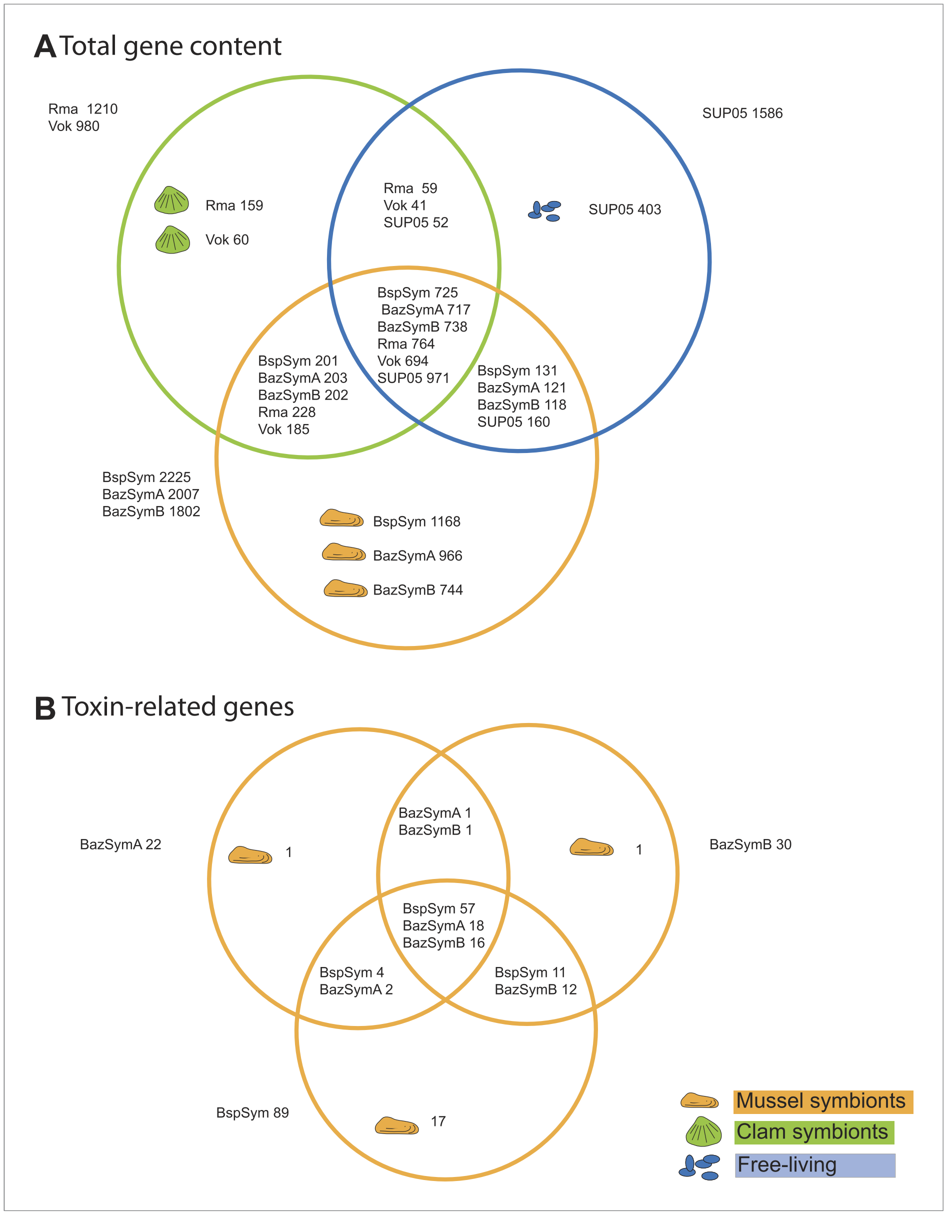

Figure 2. Genes shared between the *Bathymodiolus* and vesicomyid SOX symbionts and free-living SUP05. Protein-coding sequences from the *Bathymodiolus* sulfur-oxidizing (SOX) symbiont were compared to the clam symbiont genomes and to the SUP05 metagenome from ***Walsh et al. (2009)*** with BLAST score ratios (BSR). (**A**) Venn diagram of the shared and unique gene content in the clam symbionts, mussel symbionts, and SUP05 bacteria. Predicted protein sequences of each mussel symbiont were compared to a combined data set of the clam symbionts (Rma and Vok) and SUP05. Similarly, protein sequences of each clam symbiont were compared to a combined data set of mussel symbionts (BspSym, BazSymB, and BazSymA). Depending on the reference genome, the number of shared genes varies slightly and possibly reflects the presence of paralogous genes and redundant sequence information in these draft genomes. Abbreviations are explained in detail in ***Table 1***. The BLAST score ratio (BSR) threshold was 0.4. (**B**) Venn diagram of mussel symbiont toxin-related genes (TRGs), calculated with a BSR threshold of 0.2.

DOI: 10.7554/eLife.07966.005

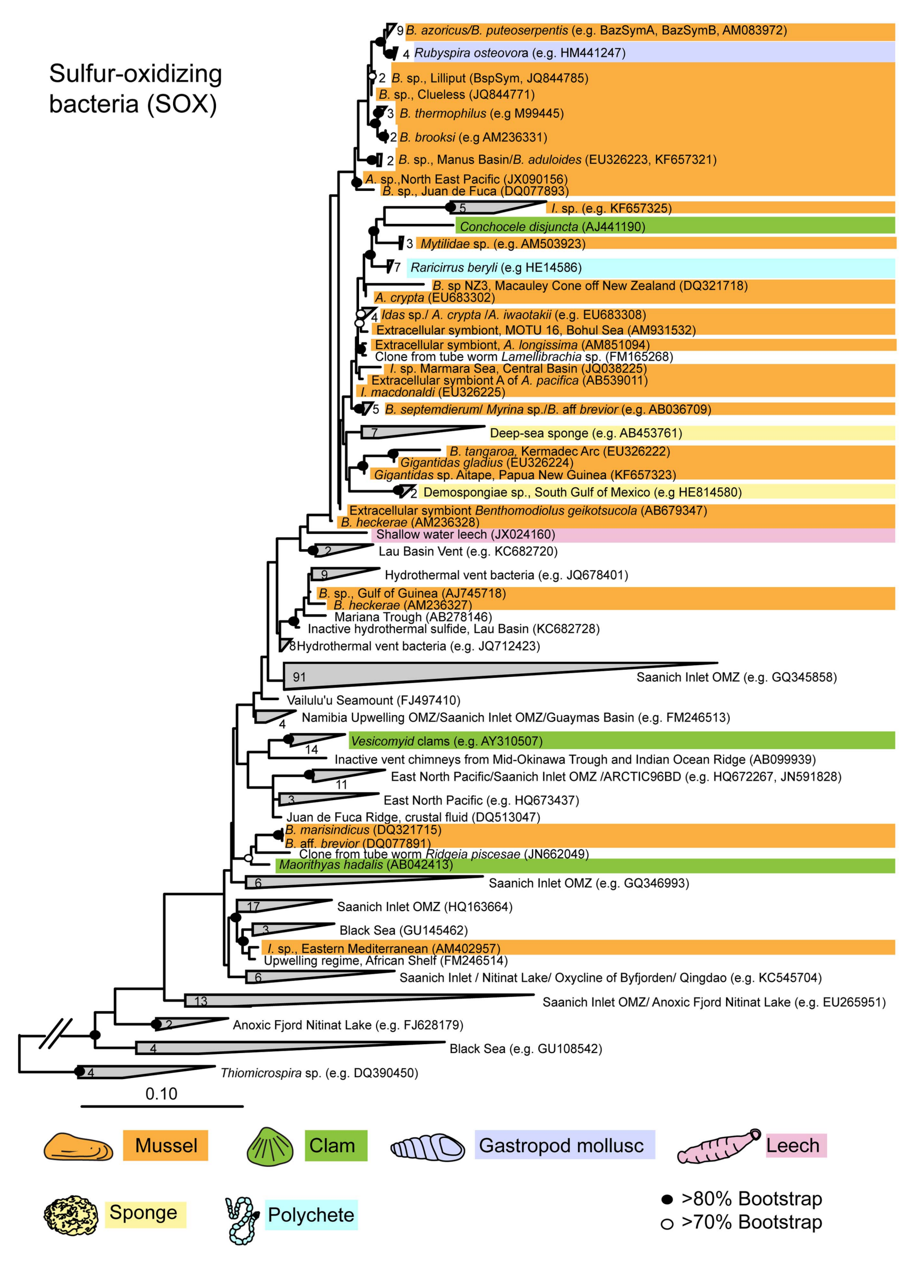

Figure 2—figure supplement 1. Maximum likelihood 16S rRNA phylogeny of the close relatives of the *Bathymodiolus* SOX symbionts. The tree was estimated from an alignment of 1653 nucleotide positions and was rooted with four sequences from *Thiomicrospira* species. The number of sequences per collapsed group is shown next to the gray blocks. Diagonal lines in the out-group branch indicate that the branch is not to scale. *B.* = *Bathymodiolus*; *A.* = *Adipicola; I.* = *Idas.*

DOI: 10.7554/eLife.07966.006

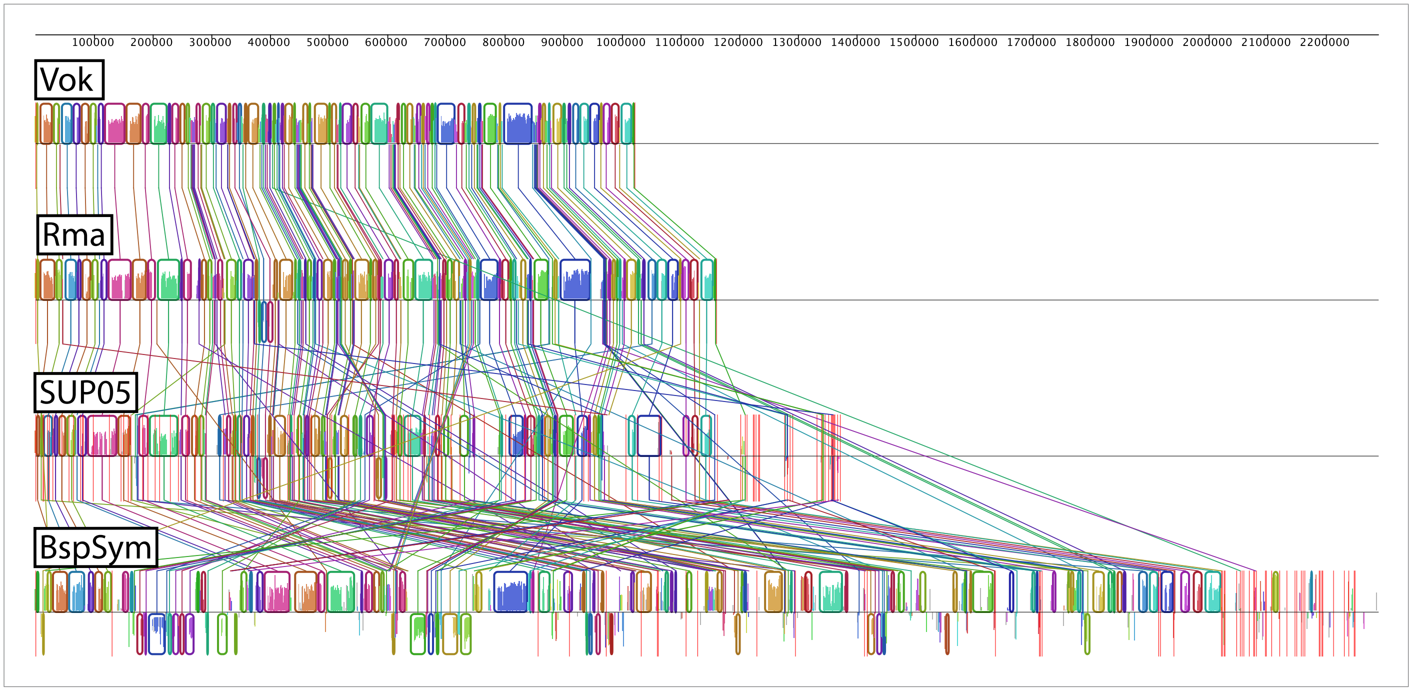

Figure 2—figure supplement 2. Whole genome alignment. Each colored block is a region of the genome that aligned to part of another genome because it is homologous and the genes are arranged in the same order. Lines crossing represent conflicting information when compared to other genomes. These are the sites where lack of synteny was observed. Red vertical lines represent contig boundaries. BspSym = SOX symbiont of *Bathymodiolus* sp., Vok = SOX symbiont *Candidatus Vesicomyosocious okutanii*, Rma = SOX symbiont of *Calyptogena magnifica* (*Ca.* Ruthia magnifica), SUP05 = free-living marine sulfur oxidizers.

DOI: 10.7554/eLife.07966.007

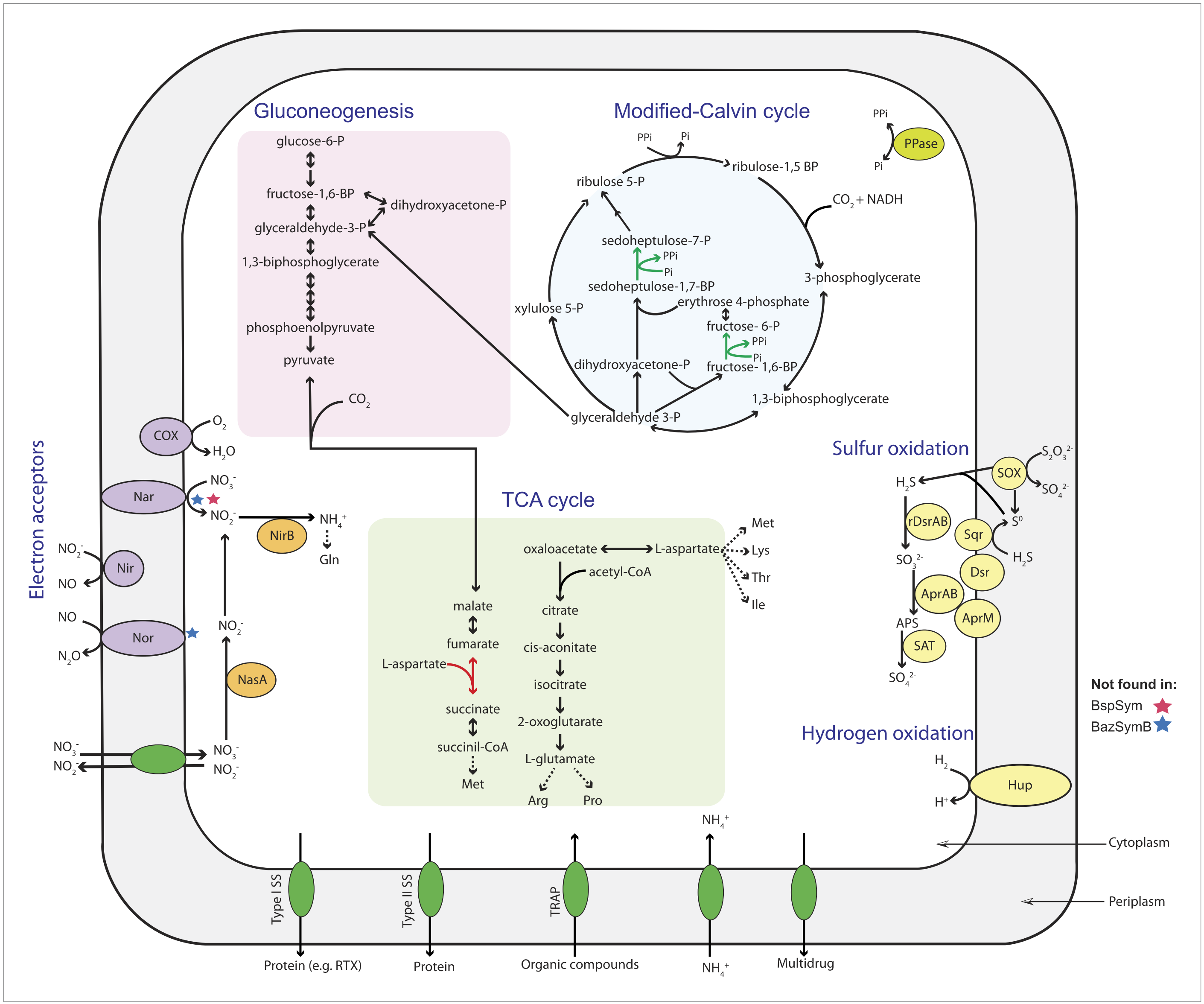

Figure 2—figure supplement 3. Metabolic reconstruction of the Bathymodiolus symbiont. Key metabolic pathways were inferred from genomic information using Pathway Tools (***Caspi et al., 2014***). Red stars indicate that the gene was not found in the *B.* sp. symbiont genome, and blue stars indicate that the gene was not found in BazSymB, but was found in BazSymA, both symbionts of *B. azoricus*. Red arrow indicates a missing enzyme that could be replaced with an alternative reaction. Green arrow indicates an inorganic pyrophosphate-dependent step in the modified version of the Calvin cycle. Nar = nitrate reductase; Nir = nitrite reductase; Nor = nitric oxide reductase; Hup = membrane-bound hydrogenase; SOX = sulfur oxidation; rDSR = reverse dissimilatory sulfite reductase; Sqr = sulfide-quinone reductase; Apr = adenylsulfate reductase; SAT = sulfate adenyltransferase; P = phosphate; BP = biphosphate; COX = cytochrome c oxidase; Gln = glutamine; Arg = arginine; Pro = proline; Met = methionine; Lys = lysine; Thr = threonine; Ile = isoleucine; PPi = inorganic pyrophosphate; PPase = soluble pyrophosphatase; SS = secretion system.

DOI: 10.7554/eLife.07966.008

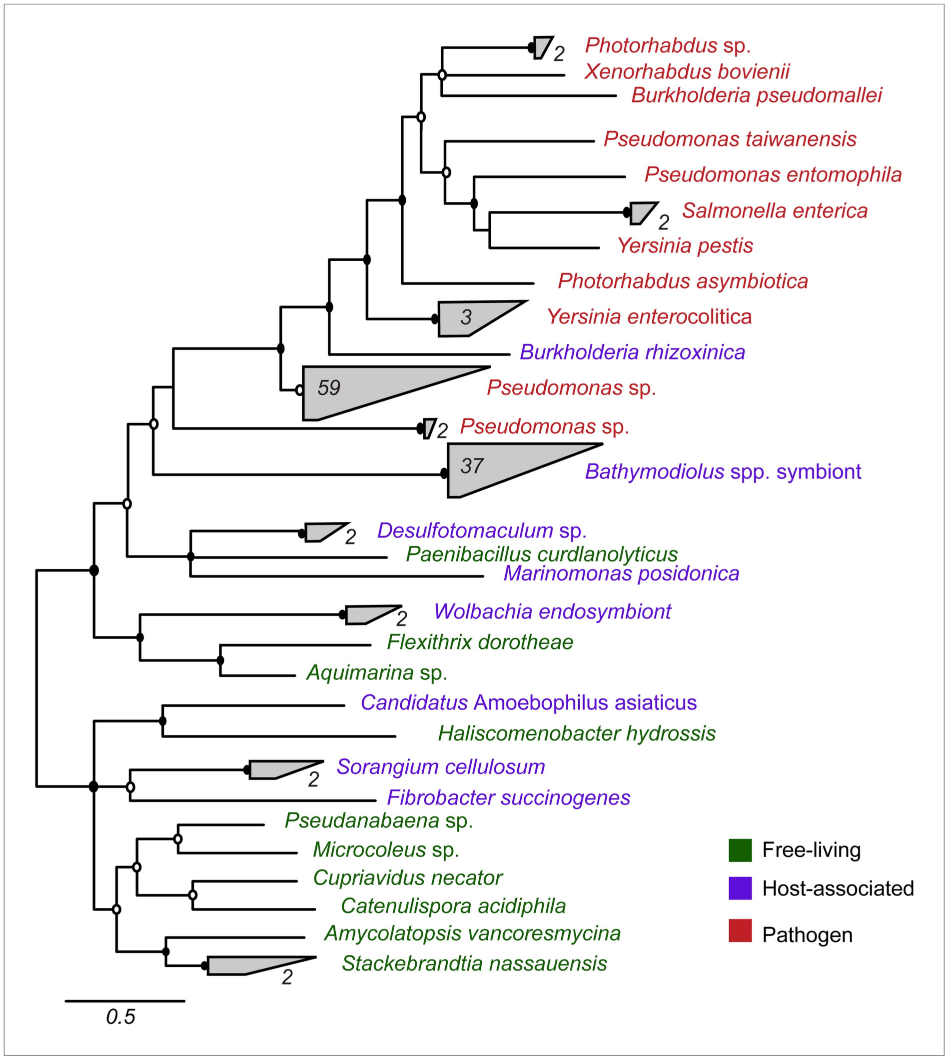

Figure 3. Phylogeny of YD repeat-containing proteins. The tree is a consensus of bayesian and maximum likelihood analyses, result of an alignment of 536 amino acids. Black circles represent branches with posterior probability >0.8 and bootstrap value >80. White circles represent branches with either posterior probability >0.8 or bootstrap value >80. The number of sequences per collapsed group is shown next to the gray bloks. Purple: organism found in intestinal microflora or in close association with another organism; green: free-living; red: pathogen.

DOI: 10.7554/eLife.07966.009

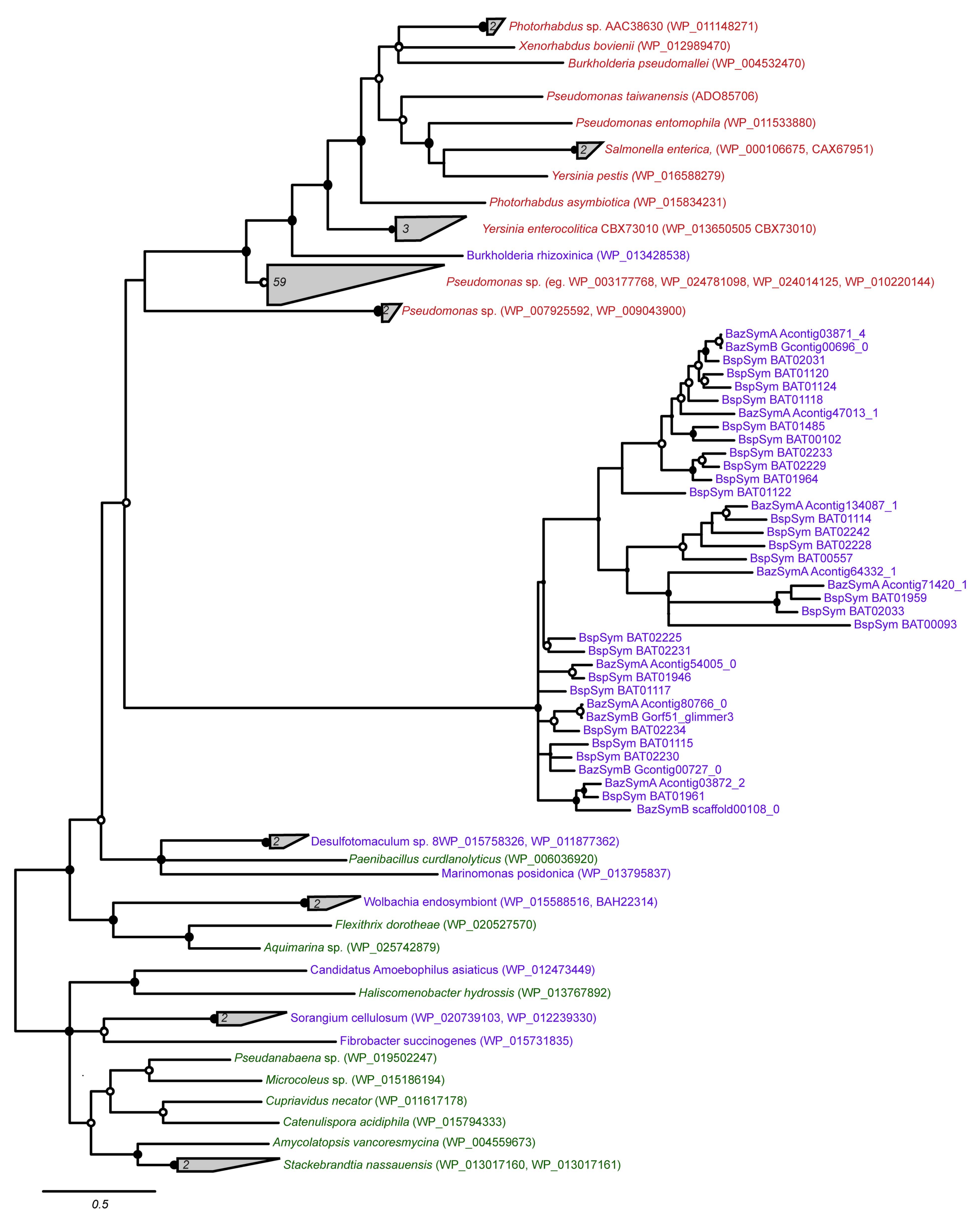

Figure 3—figure supplement 1. Consensus of bayesian and maximum likelihood phylogeny of YD proteins with identifiers. Trees were estimated from an alignment of 536 amino acids. Circles represent branches with posterior probability higher than 0.8 and bootstrap values higher than 80/100. If both reconstruction methods are significant, the circle is black, otherwise it is white. Purple: found in intestinal microflora or in close association with other

Figure 3—figure supplement 1. continued on next page

Figure 3—figure supplement 1. Continued

organisms; green: free-living; red: pathogen.

DOI: 10.7554/eLife.07966.010

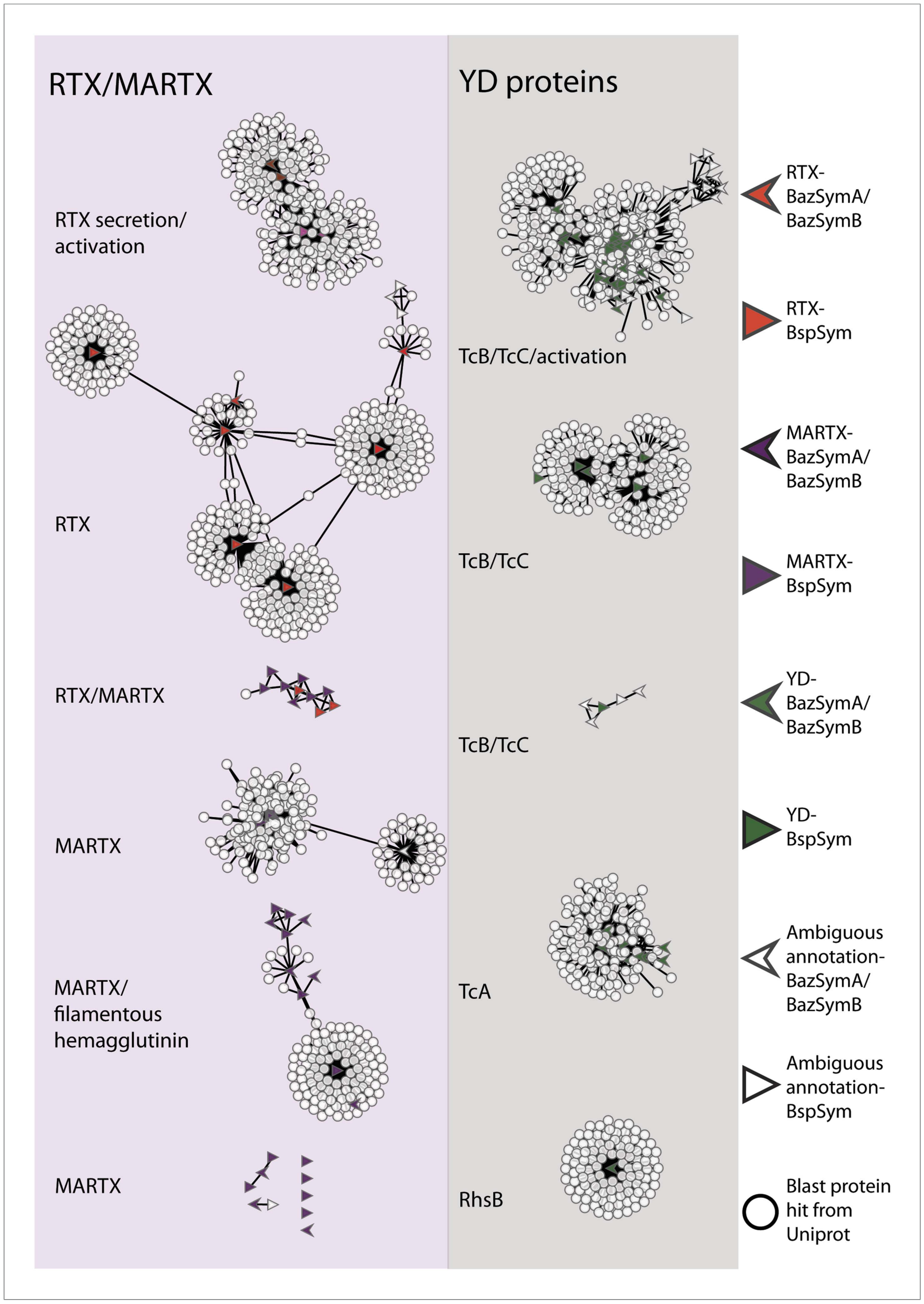

Figure 4. Protein similarity network of toxin-related proteins in the *Bathymodiolus* symbionts. Each node corresponds to a protein sequence and the links between nodes represent BLAST hits. The length of the edges is inversely proportional to the sequence similarity. Protein clusters containing RTX or multifunctional autoprocessing RTX (MARTX) proteins are shown in the red panel on the left, and sequence clusters containing YD repeats are shown in the gray panel on the right. Arrowheads are proteins from *B. azoricus* symbionts, and triangles are proteins from *B.* sp. symbionts. The symbols are colored in green if they were identified in the Bathymodiolus symbionts as YD repeat-containing genes, red if they were identified as RTX genes, and purple for MARTX genes. Some protein sequences were similar to the TRGs but not annotated as such as these are partial genes that did not have any conserved domain. If the clusters contained mostly genes with a particular annotation, we named the clusters after these annotations, for example, cluster 'TcB/TcC' contained proteins annotated as TcB or TcC.

DOI: 10.7554/eLife.07966.011

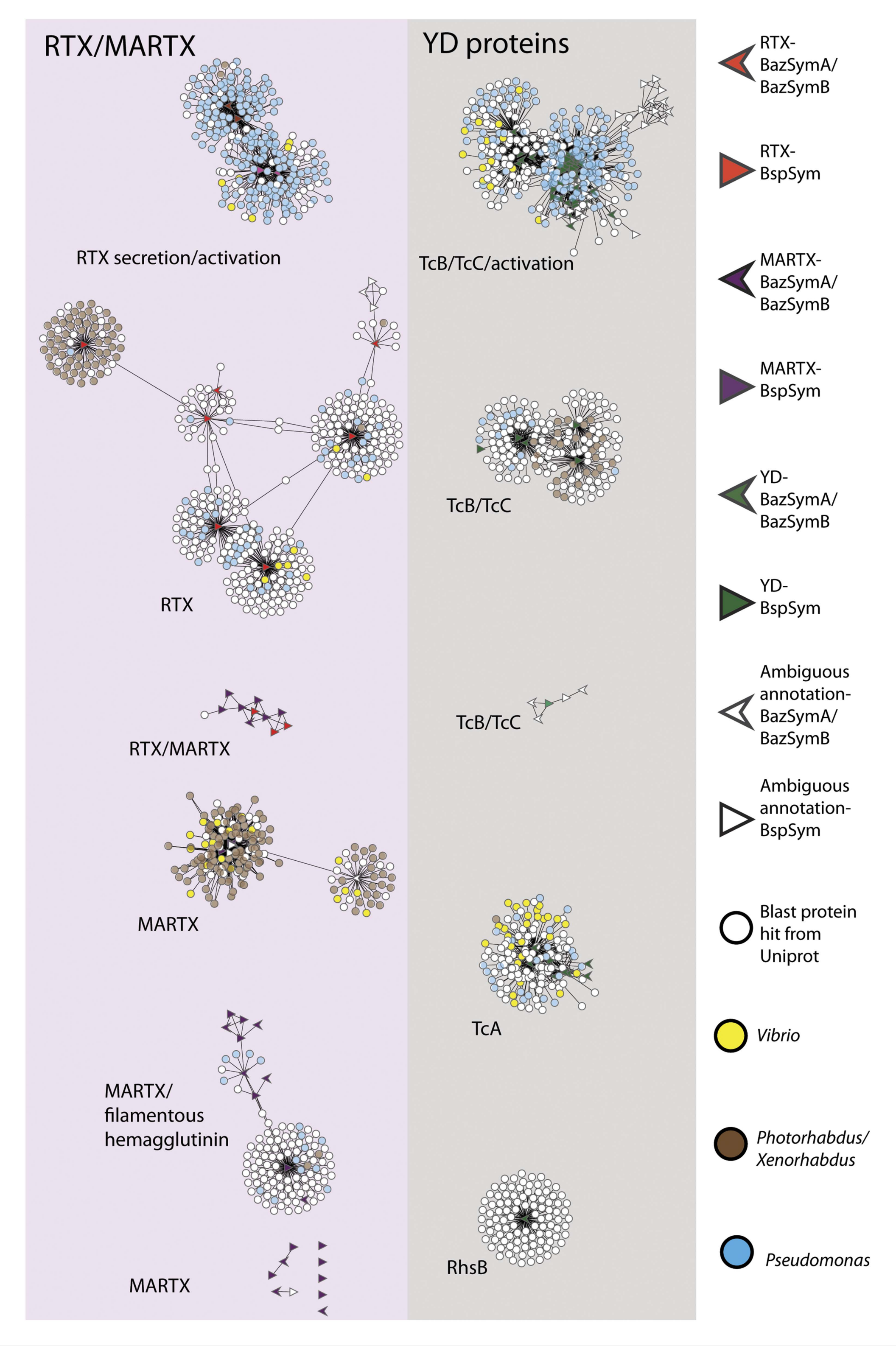

Figure 4—figure supplement 1. Network of toxin-related proteins in the *Bathymodiolus* symbionts with BLAST hits from *Vibrio*, *Photorhabdus*, *Xenorhabdus*, and *Pseudomonas* highlighted. Each node corresponds to a protein sequence and the links between nodes represent BLAST hits. The length of the link is proportional to the sequence

Figure 4—figure supplement 1. continued on next page

Figure 4—figure supplement 1. Continued

similarity. Protein clusters containing RTX or MARTX are shown in the red panel on the left. Sequence clusters containing YD repeats are shown in the gray panel on the right. Arrowheads are proteins from *B. azoricus* symbionts, and triangles are proteins from *B*. sp. symbionts. The symbols are colored in green if they could be identified in the Bathymodiolus symbionts as YD repeat-containing proteins, red if they could be identified as RTX proteins, and purple for MARTX. If the clusters contained mostly proteins with a particular annotation, we named the clusters after these annotations, for example, cluster 'TcB/TcC' contained proteins annotated as TcB or TcC.

DOI: 10.7554/eLife.07966.012

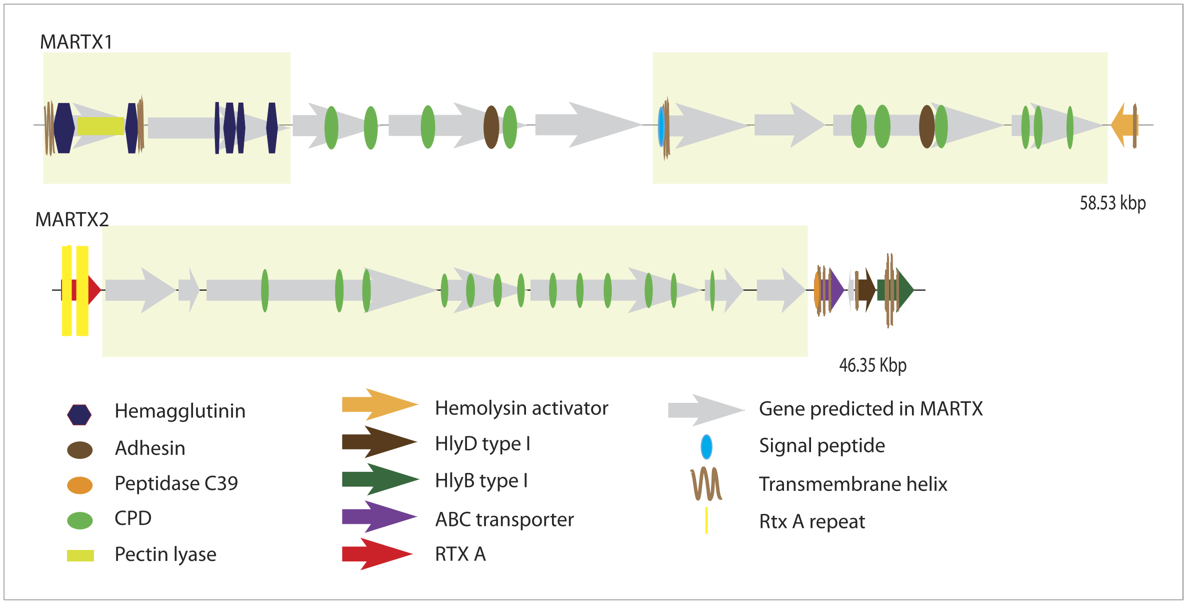

Figure 4—figure supplement 2. Genomic architecture of MARTX regions. The two MARTX regions in BspSym are shown. Operons identified by assembling transcriptome data are indicated in yellow boxes.

DOI: 10.7554/eLife.07966.013

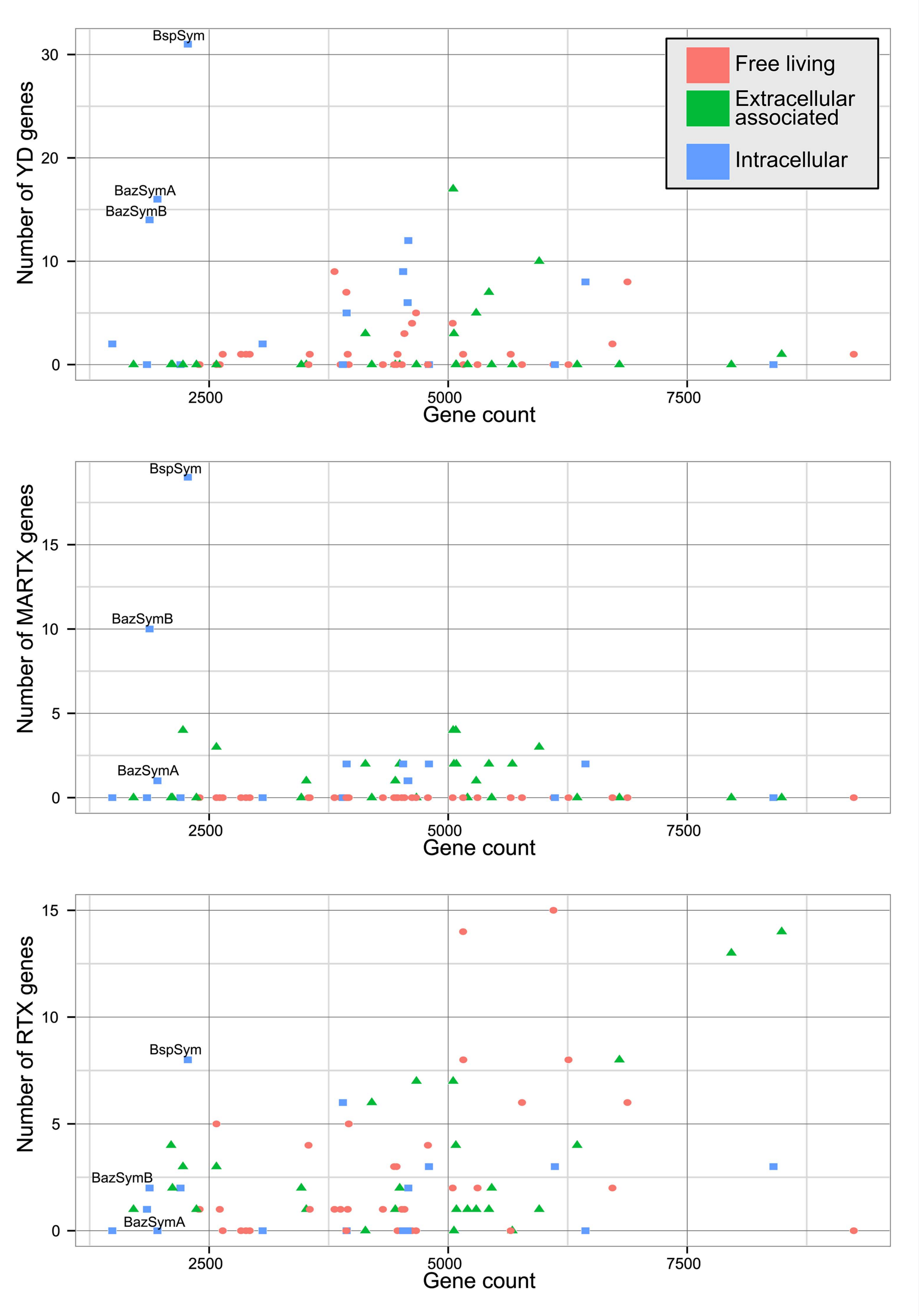

Figure 5. Distribution of the three major TRGs classes according to lifestyle. Each dot represents one sequenced genome. The sum of TRGs is on the Y axis, and the total number of genes predicted in each genome is on the X axis. Free-living bacteria are shown in red, host-associated bacteria that live outside of host cells are in green, and host-associated bacteria that can live inside host cells are shown in blue. The positions of the Bathymodiolus SOX symbionts are indicated. A detailed overview of all organisms that had similar TRGs to the SOX symbiont with the number of TRGs is shown in ***Supplementary file 1B***.

DOI: 10.7554/eLife.07966.014

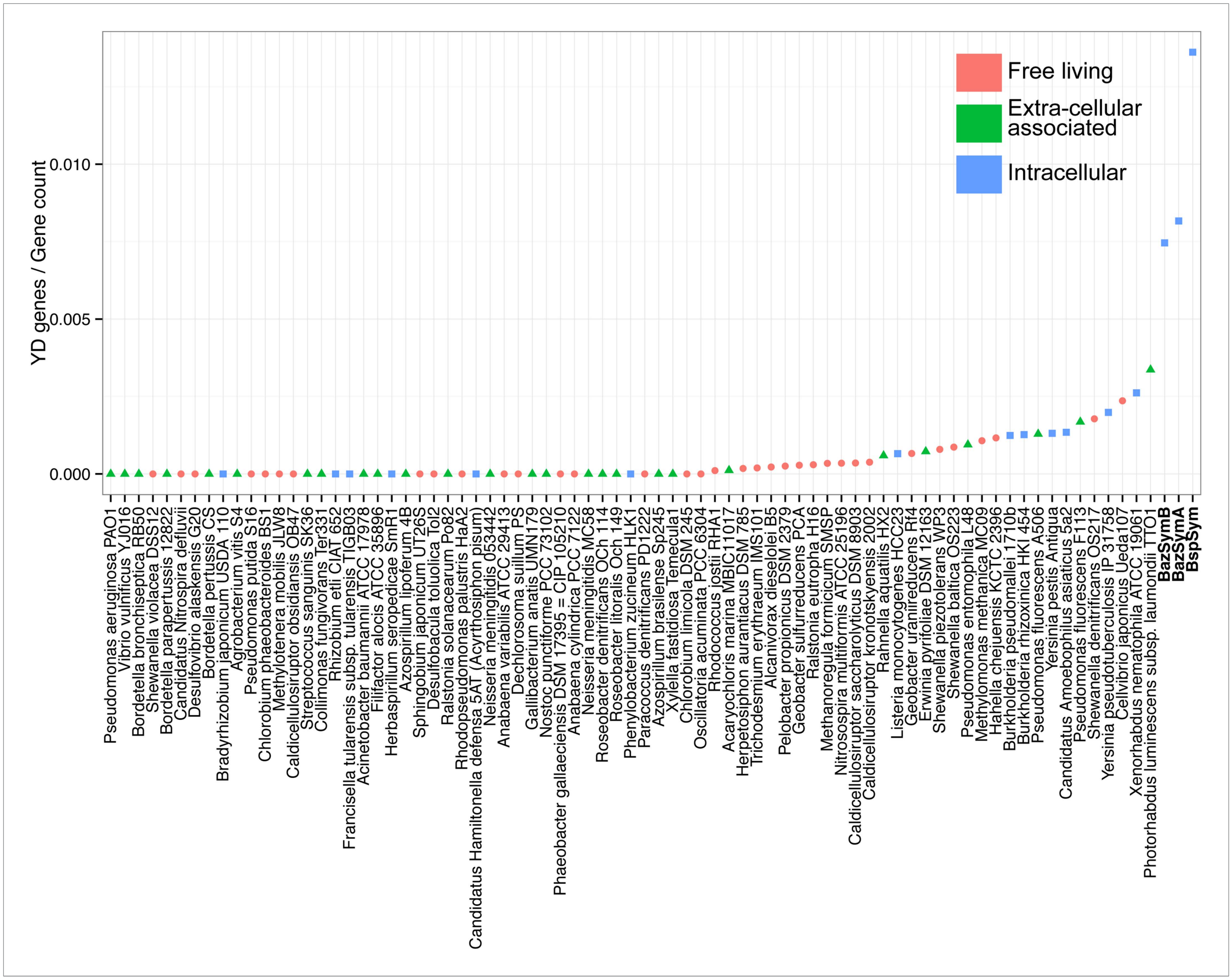

Figure 5—figure supplement 1. YD genes per genome, normalized to the total gene count. Each dot is colored by the category to which they belong. *Bathymodiolus* SOX symbionts are highlighted.

DOI: 10.7554/eLife.07966.015

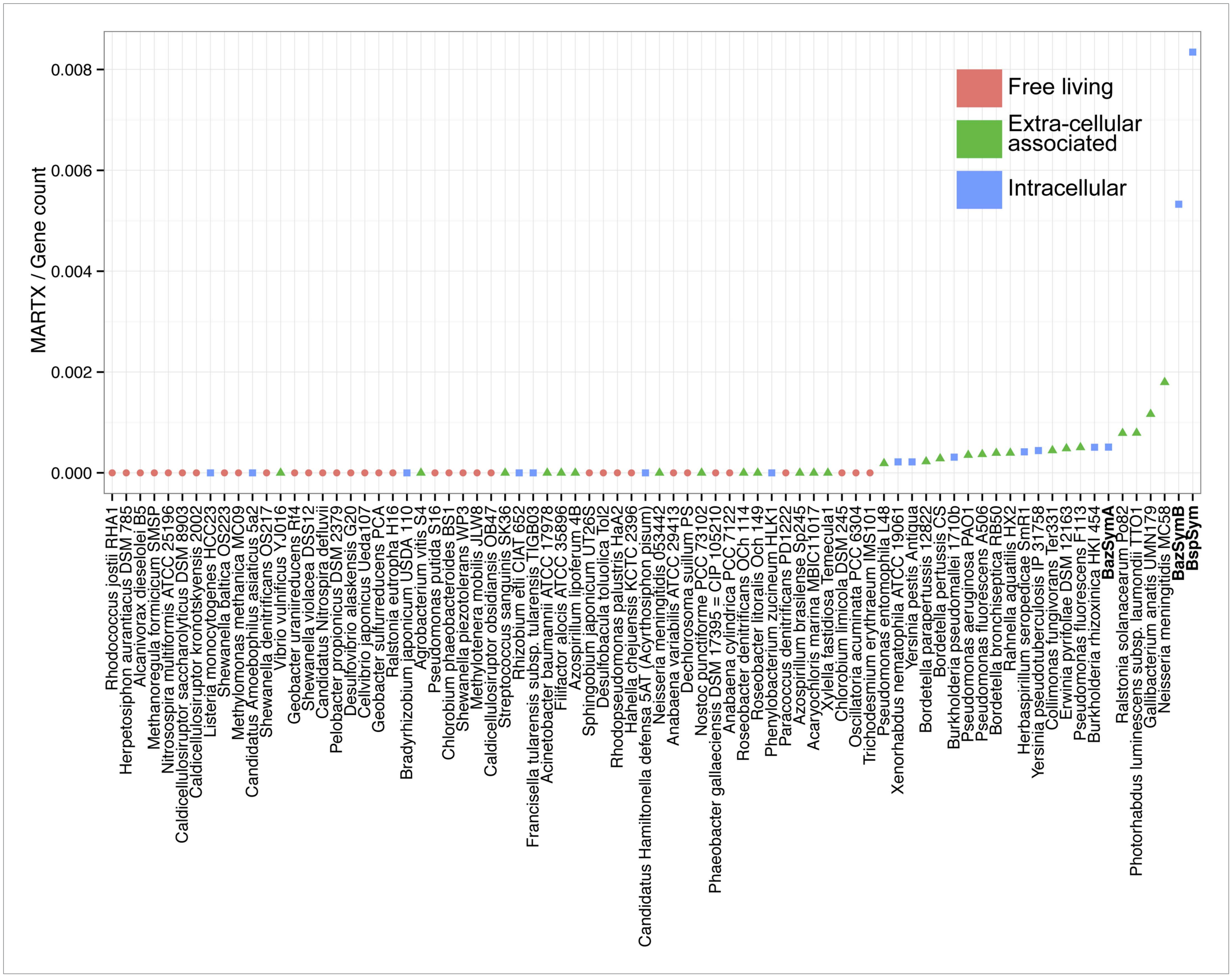

Figure 5—figure supplement 2. MARTX genes per genome, normalized to the total gene count. Each dot is colored by the category to which they belong. *Bathymodiolus* SOX symbionts are highlighted.

DOI: 10.7554/eLife.07966.016

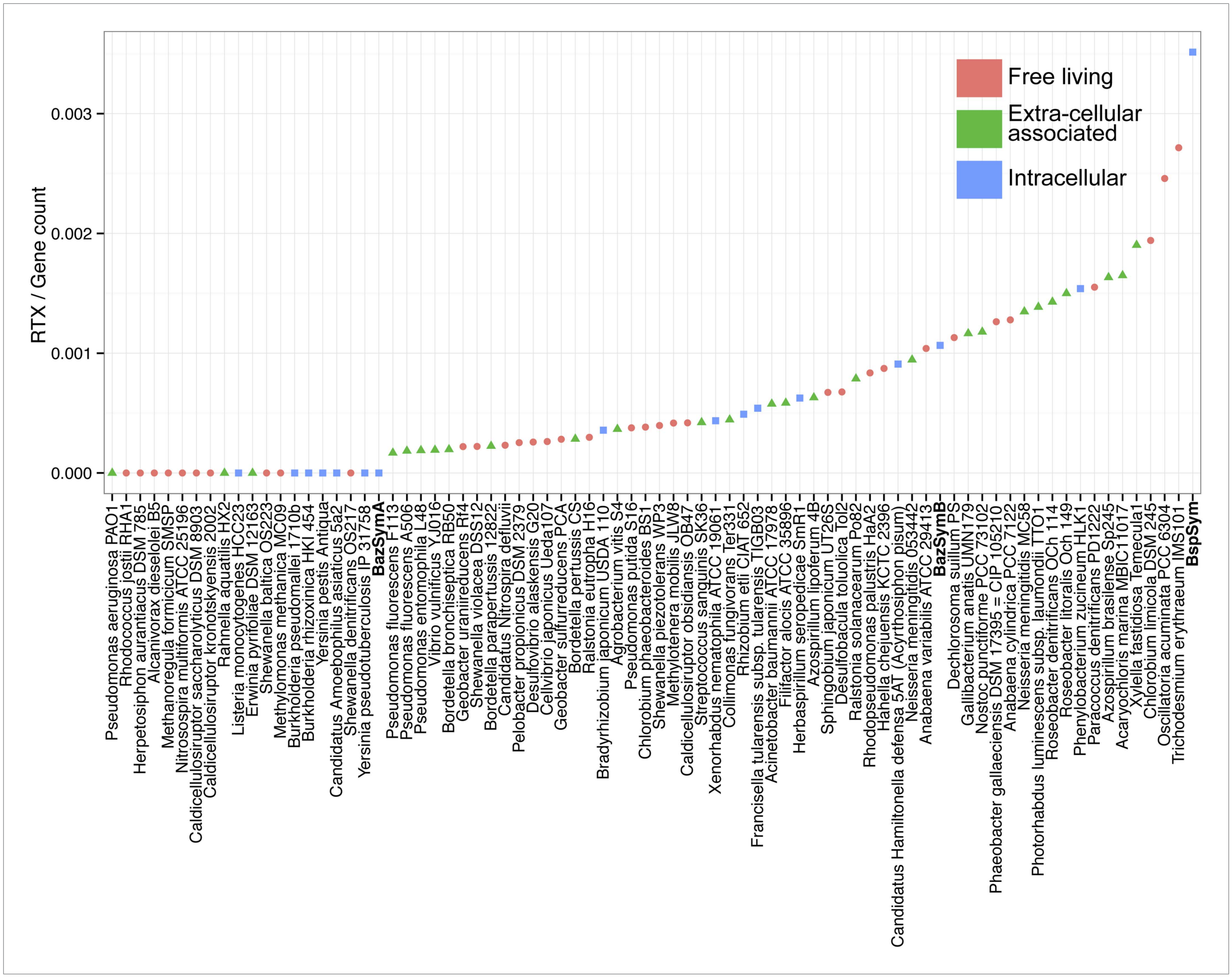

Figure 5—figure supplement 3. RTX genes per genome, normalized to the total gene count. Each dot is colored by the category to which they belong. *Bathymodiolus* SOX symbionts are highlighted.

DOI: 10.7554/eLife.07966.017

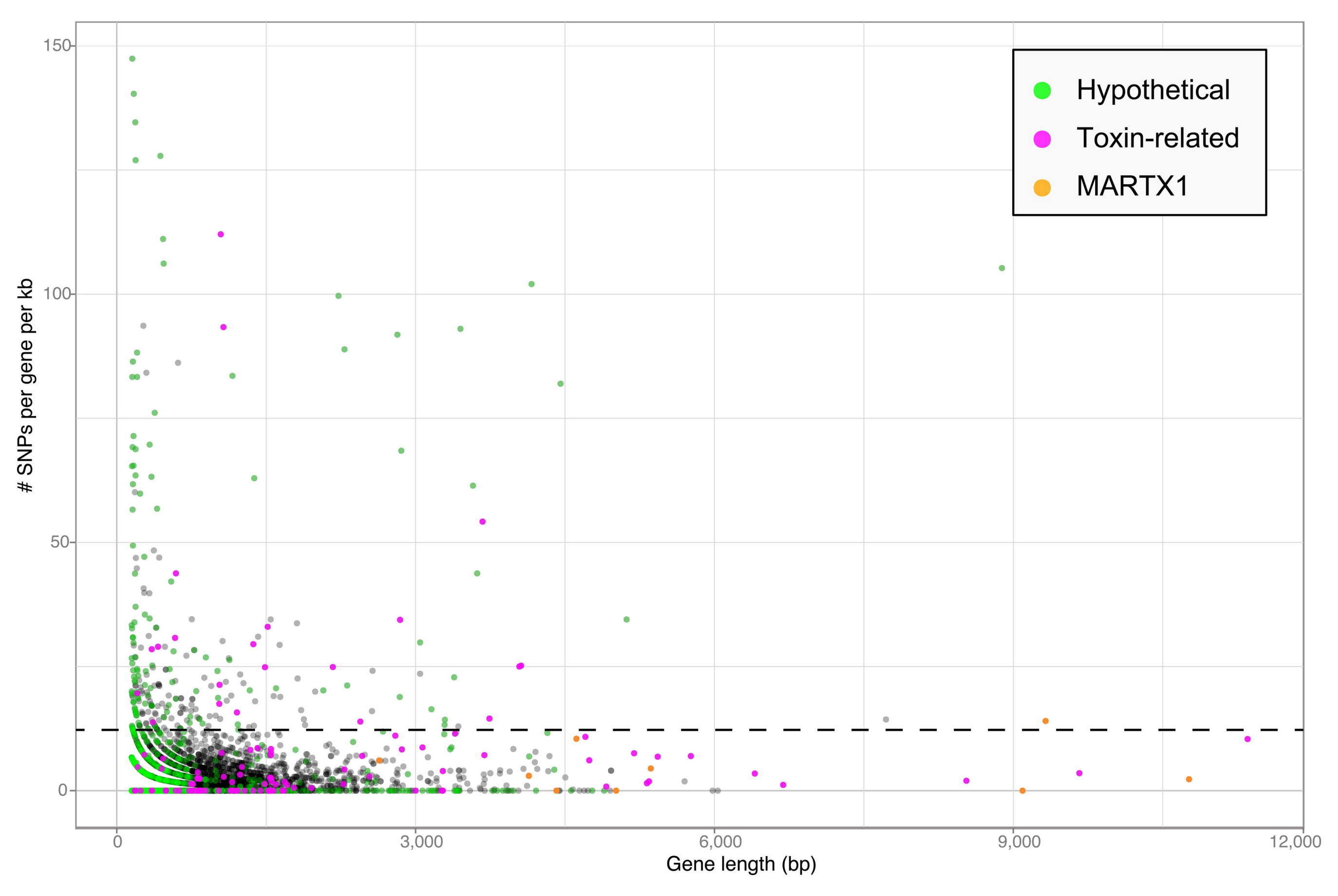

Figure 6. Single nucleotide polymorphisms per gene. The number of single nucleotide polymorphisms (SNPs) per gene was normalized according to the length minus regions of unknown sequence for genes containing N's. Genes smaller than 150 bp were not considered. The dotted line represents the median plus one standard deviation of the number of SNPs per gene per kb.

DOI: 10.7554/eLife.07966.020